Detection of Negative Emotions and High-Arousal Negative-Valence States on the Move

Valentina Markova
Dept. of Communication Engineering and Technologies Technical University of Varna Varna, Bulgaria
vim@ieee.org

Todor Ganchev
Dept. of Computer Science and Engineering Technical University of Varna
Varna, Bulgaria
tganchev@ieee.org

Kalin Kalinkov
Dept. of Communication Engineering and Technologies Technical University of Varna Varna, Bulgaria
kalin.kalinkov@mail.bg

***Abstract*— We present the overall design and implementation of a wearable system that is capable to continuously monitor and register negative emotions, high levels of emotion arousal, and high-arousal-negative-valence states from physiological signals, such as skin conductivity and ECG. This system builds on the client-server architecture. The commercially available wireless data acquisition devices Shimmer3 GSR+ and Shimmer3 ECG are used for the acquisition of physiological signals, which are then transmitted over a Bluetooth channel to a mobile phone. The mobile phone hosts the user interface and implements the data aggregation and transmission to the server, which carries all signal processing and classification tasks. Purposely developed software implements all data processing and recognition tasks on the server side and the user interface and the data communication on the client side. We report evaluation results for various setups of the binary detectors of negative emotions, high level of emotional arousal, and high-arousal-negative-valence states.**

Keywords— stress detection, GSR, ECG, wearable wireless sensors, SVM.

I. INTRODUCTION

Emotions are often assessed through analysis of verbal or non-verbal emotional expressions, or/and physiological signals. There are numerous studies on emotion assessment based on the observation, analysis, and interpretation of verbal and non-verbal behavior messages, such as facial expressions, gestures, and speech [1][2]. Recent technological advances motivated multi-modal emotion recognition that involves the combination of data streams from various physiological sensors. In this regard, classification of four emotion categories (positive/high arousal, negative/high arousal, negative/low arousal, and positive/low arousal) was performed in [3] by acquiring respiration changes, electrocardiography (ECG), electromyography (EMG) and galvanic skin response (GSR) signals. As a result of joint projects, an appropriate resources were created and released for public use [4][5]. In particular, the ASCERTAIN dataset [4] contains recordings of emotional reaction and personality traits assessment of 58 participants. It comprises data patterns extracted from ECG, GSR, Electroencephalogram (EEG), and facial activity signals evoked by 36 video clips. DEAP [5] is multimodal database for analysis of human emotional states distributed in the four quadrants of the valence-arousal space. It contains EEG and peripheral physiological sensors data obtained by 32 participants while watching 40 video clips.

The assessment of negative emotional states and conditions linked to stress has an important social and economic dimensions. These are particularly important for high-risk professions (rescue crews, firefighters, pilots, military etc.), who experience the severe consequences of continuous exposure to extreme stress. In brief, stress is defined as physiological reaction of the organism to given emotional and/or physical event. It is well known that prolonged exposure to negative emotions can negatively affect work performance, attitude, and decision making. In some cases these lead to burnout, behavioral disorders, and various mental and heart diseases. Therefore, it is very important people to be aware that they are exposed to a stressful situation, so that they will be able to take respective actions to cope with it.

Various sensor systems (mobile, wearable, wireless) for stress detection were reported [6-10]. Wireless sensor network for synchronized monitoring of a group of subjects was proposed in [7], where the system collected hearth rate (HR) data from the military trainees during several days of training in order to measure HR variability (HRV) changes before, during, and after stress exposure. In [8] a data stream from an accelerometer, ECG and GSR sensors were collected for three physical activity conditions (sitting, standing, or walking) and with three experimental settings (a baseline measurement with no stressor, measurement during the mental tasks, and a recovery segment). This study reported 92.4% accuracy of mental stress classification for 10-fold cross-validation. A real-time mobile stress recognition system that used ECG and electrodermal activity (EDA) data was proposed in [9]. The system provided instant feedback to the user and offered the possibility for off-line review of the past stressful situations by audio clips recorded at the time of the stress event. The classification accuracy of stress using only ECG data with 10-fold cross-validation and the SVM classifier was 67.5%, while accuracy of stress detection obtained using EDA pattern was 79.7%. Reportedly, the combined use of features extracted from both signals improved the classification accuracy to 82.9%. In a recent study [10], a PPG-based sensor implemented as a smartwatch device was used for continuous stress-level monitoring throughout the day. The system permits to the user to monitor alterations in his/her stress levels and provides feedback to help keeping the normal health status by avoiding high levels of stress.

978-1-5386-5558-0/18/$31.00 ©2018 IEEE

The present work is aligned along a similar idea – we designed a system for continuous monitoring, detection, and logging of physiological conditions associated with high levels of emotional arousal, negative emotions, and high-arousal-negative-valence (HANV) states. For that purpose, we integrated a client-server based system, which incorporates wearable sensors, a smart phone, and a remote server. In Section II, we present the overall system design and in Section III provide details on the signals, features, and detectors used for the detection of emotional arousal, negative emotions, and HANV states. The experimental setup and the resources used in the validation experiment are outlined in Section IV. Next, in Section V we discuss the experimental results. Finally, we conclude this work in Section VI.

II. Overall system architecture

People repeatedly experience conditions associated with high levels of stress and negative emotional states, while being at work, on the move, and during daily routine activities. Considering the need of a portable device that is capable to continuously register, evaluate, and log such conditions, we integrated a system consisting of multiple hardware and software components, some of which are wearable and other remote to the user (*cf.* Fig.1). Specifically, the major hardware components are (i) two small wearable sensors that capture ECG and GSR signals, (ii) a smart phone that hosts the user interface and serves as a transmission gateway between the wearable sensors and Internet, and (iii) a remote server where the data processing and recognition is implemented.

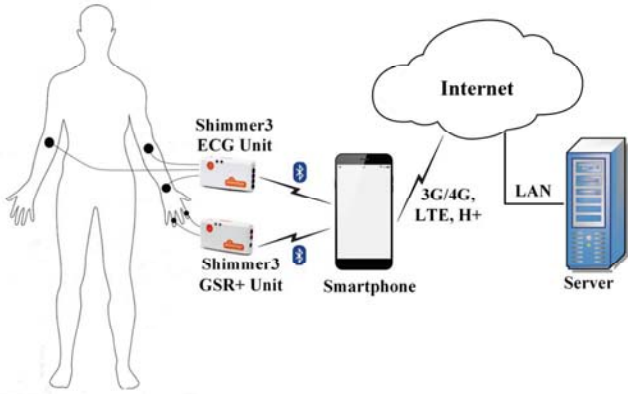

Fig. 1. Overall architecture of our system for the detection of negative emotions, high emotional arousal, and high-arousal-negative-valence states.

In brief, the physiological signal acquisition was implemented my means of the commercially available[1] wearable sensors Shimmer3 GSR+ and Shimmer3 ECG, which are conveniently attached on the wrist like a watch. Specifically, we rely on the Lead I ECG signal acquisition scheme with three disposable Ag-AgCl electrodes attached close to the cubital fossa on the right and left forearms, and on the right-arm wrist. The skin conductance signal was obtained between the index and ring fingers with reusable Ag-AgCl electrodes, attached on the palmar surface of the intermediate phalanges. Such an attachment scheme provide steady connections of the electrodes and was found less discomforting.

The control of the Shimmer3 units and the data transmission over the Bluetooth channel are carried out with a purposely developed application which makes use of the publicly available Shimmer API. In such a way, the physiological signals collected through the Shimmer units are temporary stored to the smartphone in CSV format and are then uploaded to a remote server running OS Ubuntu 16, where the signal processing and classification tasks are carried out.

The smartphone application can visualize some descriptors such as the hearth rate (HR), hearth rate variability (HRV) computed based on the ECG signal, the direction of their change, and color-coded information about the presence/absence of negative emotions and the HANV state, as well as the raw ECG and GSR signals.

In the following sections we focus on the signal processing and classification methods incorporated in these detection tasks.

III. Detection framework

Once the ECG and GSR signals are delivered to the server, these are subject to signal parameterization and classification (*cf.* Fig. 2). The signal parameterization process consists of signal pre-processing, feature extraction, feature post-processing, and composing of the feature vectors for the Valence, Arousal and HANV classifiers. These classifiers are designed as detectors, which recognize negative emotional states, high level of emotional arousal, and HANV states. The outcome of the classification process is a binary decision about the label of an input recordings, where the output of the valence binary detector is either "*Negative Emotion*" or "*Other*"; the output of the high-arousal binary detector is either "*Low Arousal*" or "*High Arousal*"; and the output of the HANV detector is either "*HANV*" or "*Other*".

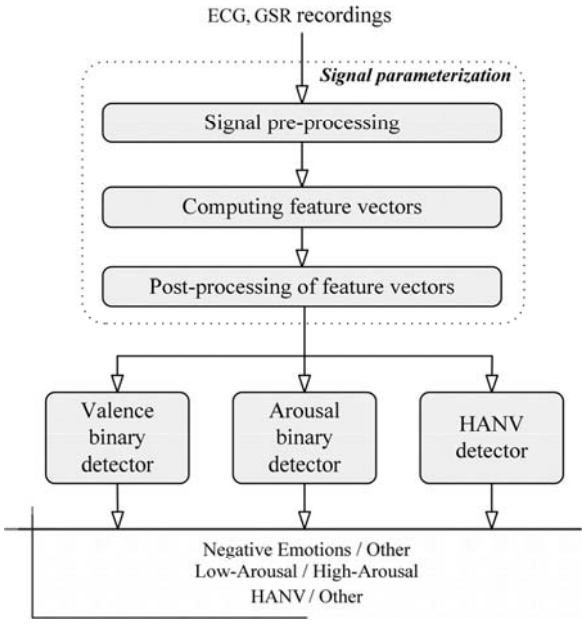

Fig. 2. Block diagram of the affect/stress recognition system

[1] The Shimmer3 GSR+ and Shimmer ECG units are products of the Shimmer company (http://www.shimmersensing.com/)

In Fig. 3, we summarize the feature extraction steps for the GSR- and ECG-based features -- a detailed description is offered in Sections III.A and III.B, respectively.

A. GSR-based features

In particular, we consider extraction of descriptors which characterize the two informative components of the GSR signal: skin conductance level (SCL) and skin conductance response (SCR). The SCL component is the slowly varying tonic level, which changes slightly within tens of seconds to minutes. The tonic level usually differs among people. The SCR component, also known as *GSR peaks*, is the elicited reaction to specific stimuli, which are informative for acute stress level assessment. For the purpose of arousal detection and HANV detection, we estimate both the GSR peaks and the mean and standard deviation of the SCR following the processing steps shown in Fig. 3. In particular, we first convert the GSR signal to skin conductivity signal and apply a low-pass filter (LPF) with cut-off frequency 0.5 Hz in order to eliminate interferences due by body movement or poor contact between skin and electrodes. Subsequently, the signal obtained to this end is down-sampled to 25 samples per second. A median filter is applied on the output signal.

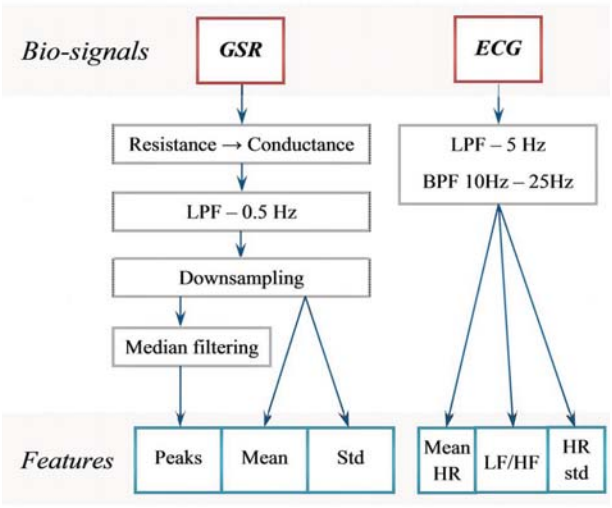

Fig.3. Block diagram of the ECG- and GSR-based feature extraction

The SCR component is obtained after_subtraction of the median filter output from the raw signal. Eventually, we count the number of GSR peaks for each one-minute segment corresponding to a single audio-visual stimuli and compute the mean and the standard deviation of the tonic level. Making use of the features obtained to this end, we computed additional attributes (*cf.* Table I) based on the statistics of individual features, their first and second differences, and the properties of their distributions.

B. ECG-based features

The ECG signal is subject to filtering with a pair of non-orthogonal filters: a LPF with cut-off frequency 5 Hz and a band-pass filter with cut-offs 10 Hz and 25 Hz. Both filters have suppression of −40 dB in the stopband. Hearth rate (HR) analysis is performed for the low-pass filtered signal, and next we compute the parameters HR mean and HR standard deviation, which are considered representative descriptors of the ECG signal. Next, a frequency-domain analysis is carried out and based on the power spectrum we estimate the ratio LF/HF between the energy levels in the low-frequency range (LF ϵ [0.04, 0.15] Hz) and the high-frequency range (HF ϵ [0.15, 0.4] Hz). In addition, we compute the power in the very low frequency range (VLF ϵ [0, 0.04] Hz), the normalized values of the VLF, LF, and HF in percentages, and statistics over the RR and the NN intervals (*cf.* Table II).

TABLE I. LIST OF GSR-BASED FEATURES

ID	Designation	GSR-based Feature Description
1	Num_Peaks	Number of peaks for the whole record
2	Cond_mean_all	Mean conductance of the whole record
3	Cond_max	Maximum amplitude of the peaks
4	Cond_min	Minimum amplitude of the peaks
5	Cond_mean	Mean conductance of the peaks
6	Cond_rms	RMS of the peaks conductance
7	Cond_std	Standard Deviation of the peaks conductance
8	Cond_mean_abs	Mean absolute value of the peaks conductance
9	Resist_mean	Mean resistance for the whole record
10	Skewness	Skewness of the peaks distribution
11	Kurtosis	Kurtosis of the peaks distribution
12	d_1	First Degree Difference (FDD)
13	d_2	Second Degree Difference (SDD)
14	d_1_div_std	Ratio of FDD and Standard deviation
15	d_2_div_std	Ratio of SDD and Standard deviation
16	Specpow2_4	Power in the band 0-2.4Hz

TABLE II. LIST OF ECG-BASED FEATURES

ID	Designation	ECG-based Feature Description
1	BPM	Mean Heart Rate
2	HRV	Heart Rate Variability
3	VLF	Power in the very low-frequency band (0-0.04) Hz
4	LF	Power in the low-frequency band (0.04-0.15) Hz
5	HF	Power in the high-frequency band (0.15-0.4) Hz
6	LF_nu	Power in the low-frequency band (normalized units)
7	HF_nu	Power in the high-frequency band (normalized units)
8	VLF_p	Power in the very low-frequency band (%)
9	LF_p	Power in the low-frequency band (%)
10	HF_p	Power in the high-frequency band (%)
11	mIBI	Mean inter-beat interval (Mean RR)
12	mNN	Mean NN interval
13	NN_max	Maximum length NN interval
14	NN_min	Minimum length NN interval
15	pNN50	Percent of successive NN intervals, with difference more than 50ms
16	SDNN	Standard Deviation of NN intervals
17	RMSSD	RMS of the successive difference of NN intervals
18	SD	Standard deviation of RR intervals

C. Feature post-processing

The ECG feature vectors obtained to this end are subject to dynamic range normalization. Here, assuming Gaussian distribution for all ECG features, we implemented dynamic range normalization by subtracting the mean value and dividing by the standard deviation. Next, the GSR features are scaled to the dynamic range [0, 1] by dividing with the maximum value. Finally, the post-processed features, computed for the GSR and ECG signals, are staked together to form an aggregate feature vector. The aggregate feature vectors are then fed to the person-specific detectors of negative valence, high arousal, and HANV states, which infer the presence/absence of negative emotions, high-arousal, and HANV states for the input.

D. SVM-based detectors

We trained separate detectors for high levels of emotional arousal, negative valence emotions, and HANV states for each user. All detectors make use of a support vector machine (SVM)-based classifier with polynomial kernel, which implements the *L1* soft-margin classifier trained with the Sequential Minimal Optimization (SMO) method. In all experiments we followed the leave-one-out method and fine-tuned the SVM parameters with grid-search. The search range was set as follows: *box constrain* $C \in [10^{-6}, 1]$ with step $10^{0.2}$, *tolerance* $\varepsilon \in \{10^{-8}, 10^{-7}\}$, and *polynomial order* $p \in \{1, 2, 3\}$.

IV. Experimental setup

In a comparative evaluation of different setups, we made use of the resources described in Section IV.A and the SVM-based detectors outlined in Section III.D.

In total, we carried out eighteen experiments, involving three feature subsets, evaluated for each of the three detectors considered here. These feature subsets include: (i) the raw feature vectors, (ii) a subset of features selected based on Fisher's discriminant ratio (FDR-selected), and (iii) a subset selected after applying Fisher's discriminant ratio on the post-processed feature vectors (*Norm+FDR*-selected). The feature post-processing was performed as specified in Section III.C. The *ground true* labels C_n were obtained based either on the tags of the audio-visual stimuli or on the self-reported tags specified by each user.

A. Resources

We collected a small dataset needed for the validation of the proposed conceptual design, the development of methods and models, and the evaluation of technology. For the purpose of dataset collection we made use of the Shimmer3 GSR+ and Shimmer3 ECG units attached to the forearms of the users as shown in Fig. 4.

Fig. 4. Data acquisition setup used during the data collection and technology evaluation experiments.

During the data acquisition we used audio-visual stimuli, which are a subset of the music video clips specified in the DEAP database [5]. Specifically, we made use of seventeen musical clips, sixteen of which are categorized in the four quadrants of the Valence–Arousal emotion model. One video clip was used for setting the baseline *relax condition* at the beginning of each recording session.

In total, ECG and GSR recordings of four healthy subjects (two males and two females) aged between 23 and 49 years were collected. The sampling frequency for the ECG and GSR signals were set to 512 Hz and 128 Hz, respectively. A notch filter with cut-off frequency 50 Hz was applied to reduce the power grid interferences. The signal processing and feature extraction steps follow the procedure outlined in Section III.

B. Performance Evaluation Metrics

The accuracy of the binary detectors for negative valence, high levels of emotional arousal, and HANV was computed as the weighted sum of the class-specific accuracy obtained for the two classes – *false* and *true*:

$$Accuracy = \frac{1}{2}\left(\frac{n_{cf}}{n_f} + \frac{n_{ct}}{n_t}\right), \qquad (1)$$

where n_{cf} (true negative) is the number of correctly detected instances associated with category C_n=*false*, and n_{ct} is the number of correctly detected vectors for category C_n=*true*. For each of these detectors and feature sets we carried out two experiments: (i) with C_n based on the person-independent tags specified for the audio-visual stimuli, and (ii) with labels C_n based on the self-reported values for the valence and arousal tags.

Based on the *Accuracy* (1) obtained for the individual detectors, we computed the *Average Accuracy* for all users (M=4) in percentages:

$$A_{Average} = 100.\frac{1}{M}\sum_{m=1}^{M} Accuracy(m)\,, [\%]. \qquad (2)$$

In the following discussion, we report only the *Average Accuracy* as it provides an overall estimation of the recognition performance.

V. Experimental results

Making use of the dataset and the performance metrics outlined in Section IV, we performed a comparative evaluation of eighteen detection setups. These include three different compositions of the feature vectors in the context of binary detection of high levels of emotional arousal, negative emotions, and HANV states, and two methods for setting the *ground true* labels C_n. The first assumes ground true labels based on the audio-visual stimuli tags, and the second relies on the self-reported tags of each person.

The *Average Accuracy* results for all setups are summarized in Table III. As shown in the table, the post-processing of the feature vectors for unifying their dynamic range facilitates the detection of high levels of emotional arousal for both self-reported tags and audio visual-stimuli tags.

TABLE III. AVERAGE ACCURACY OBTAINED IN THE EVALUATION SETUPS

	Self-Reported Tags			Audio-Visual Stimuli Tags		
	High Arousal	*Negative Emotions*	*HANV*	*High Arousal*	*Negative Emotions*	*HANV*
Raw features	68.9%	72.1%	70.7%	74.1%	79.9%	72.2%
FDR-selected	71.7%	77.2%	75.6%	65.6%	84.6%	74.2%
Norm+FDR-selected	71.9%	72.1%	70.7%	77.2%	81.7%	69.1%

However, the detectors of negative emotions and HANV states show their highest recognition performance for the feature subset selected with Fisher's discriminant ratio (FDR), without post-processing. We explain this with the different relative importance of GSR-based and ECG-based features for the different detection setups. The scaling of the dynamic range of GSR-based features definitely helps for the detection of high arousal, while the z-norm applied on the ECG-based features discards important information about the mean value of these features. The last seems to impede the detection of negative emotions and HANV states.

Another interesting observation is that the detector of HANV states has a slightly higher *Average Accuracy* when the ground true labels are based on the self-reported tags. This is evident for the feature subsets selected with Fisher's discriminant ratio, regardless of whether feature scaling is performed, or not. This observation indicates that some additional efforts spent by the users could improve the HANV detection performance. This effort is needed in order to fill in a questionnaire that evaluates the audio-visual stimuli along the Valence-Arousal axes. However, when this option is not available (for example, for people with special needs or for non-cooperative users), the detectors can be based on the person-independent tags assigned to the audio-visual stimuli, without significant loss of detection accuracy.

VI. CONCLUSIONS

The experimental results support that the proposed approach has the potential to provide the required functionality for the automated detection of emotional arousal, negative emotional states, and high-arousal-negative-valence conditions from physiological signals. The reported detection accuracy for a single-stimuli recording is in the range 75% – 85%, which is insufficient for real-life applications.

Future research efforts will be focused on the definition of appropriate setups, which would allow for the acquisition and use of contextual information, and on the development of advantageous signal parameterization methods, which derive signal feature vectors purposely-designed for the stress detection task.

REFERENCES

[1] M.S. Hossain, G. Muhammad, "An Emotion Recognition System for Mobile Applications", *IEEE Access*, vol.5, pp.2281-2287, 2017.

[2] Sh. Malwatkar, R. Sugandhi, A.R. Mahajan, "Human Affect Recognition System based on Survey of Recent Approaches", *Intern. Journal of Computer Applications,* vol.158, no.6, pp.10-17, 2017.

[3] J. Kim, E. Andre, "Emotion Recognition Based on Physiological Changes in Music Listening", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.30, no.12, pp.2067–2083, 2008.

[4] R. Subramanian, J. Wache, M.K. Abadi, R.L. Vieriu, S. Winkler, N. Sebe, "ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors", *IEEE Trans. on Affective Computing*, vol.9, no.2, pp. 147–160, 2018.

[5] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdan, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, "DEAP: A Database for Emotion Analysis Using Physiological Signals", *IEEE Trans. on Affective Computing*, vol.3, no.1, pp. 18–31, 2012.

[6] V. Markova, T. Ganchev, "Technological Support to Stress Level Monitoring", Book Chapter 6 in '*Enhanced Living Environments: From Models to Technologies*', IET Publisher, pp 133-160, 2017.

[7] E. Jovanov, A. Lords, Raskovic D., Cox P., Adhami R., Andrasik F., "Stress Monitoring Using a Distributed Wireless Intelligent Sensor System". *IEEE Engineering in Medicine and Biology Magazine*, June 2003, pp. 49–55.

[8] F. Sun, C. Kuo, Cheng H., Buthpitiya Sh., Collins P., Griss M., "Activity Aware Mental Stress Detection using Physiological Sensors". *Proc of the Int. Conf. on Mobile Computing, Applications, and Services*, Springer, 2010, pp. 211–230.

[9] V. Alexandratos, "Mobile Real-Time Stress Detection". MSc Thesis, Delft University of Technology, Eindhoven, The Netherlands. January 9, 2014.

[10] T. Hao, Walter, K. N., Ball, M. J., Chang, H.-Y., Sun, S., Zhu, X., "StressHacker: Towards Practical Stress Monitoring in the Wild with Smartwatches". *AMIA Annual Symposium Proceedings*, 2017. pp. 830–838.